

Comparing Chains of Order Statistics

Charles Hoffman^a and Corey Manack^b

Addresses.

^a Department of Mathematics
Finger Lakes Community College
3325 Marvin Sands Drive
Canandaigua, NY, 14424
USA.
Email: hoffmace@flcc.edu

^b Department of Mathematics
Amherst College
Box 2239, P.O. 5000
Amherst, MA 01002-5000
USA
Phone: (413) 542-2100
Fax: (413) 542-2550
Email: cmanack@amherst.edu

Abstract. Fix $0 < k \leq m \leq n$, and let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be continuous, independent, and identically distributed random variables. We derive a probability distribution that compares the performance of a k -out-of- m system to a k -out-of- n system. By virtue of uniformity, we may recast our method of comparison to enumerating lattice paths of a certain exceedance, invoking the Chung-Feller Theorem and Ballot Numbers in our derivation. Another bijection shows that our probability distribution describes the proportion of the first $2k$ steps lying above $x = 0$, for a $(m + n)$ -step integer random walk, starting at $x = 0$ and terminating at $x = m - n$.

Keywords. Lattice path enumeration, Catalan numbers, ballot numbers, Chung Feller Theorem, Reliability Theory,

AMS classification. 05A15, 05A19, 60G50.

1. INTRODUCTION

This paper establishes a new probability distribution for comparing two sets of random variables $\mathcal{X} = \{X_1, \dots, X_m\}, \mathcal{Y} = \{Y_1, \dots, Y_n\}$ (Theorem 3.7), where $\mathcal{X} \cup \mathcal{Y}$ is a set of continuous, independent, and identically distributed (henceforth abbreviated as c.i.i.d) random variables. Our comparison method is inspired by dice comparisons from the board game RISK, where two groups of dice are rolled, ranked, and top performers are then pairwise compared. The essence of our derivation will be similar, taking pairwise comparisons between the 1st, 2nd, \dots , k th smallest values from each system ($k \leq \min\{m, n\}$). Our distribution then determines, for each l between 0 and k , the probability that exactly l of the k comparisons favored system \mathcal{Y} . Thus, our probability density function consists of four integer parameters k, l, m, n . Throughout the rest of the paper, we assume $0 < k, 0 \leq l \leq k \leq m \leq n$, where the last inequality creates no loss in generality.

An immediate application is in reliability analysis. Using terminology from the field, our distribution compares the performance of a k -out-of- m system to a k -out-of- n system, when all components are c.i.i.d.. Traditionally, k -out-of- n systems are compared by comparing the distributions of the respective k th order statistics (see, e.g., [5] or [7]). To make such a comparison, the component distributions must be known. Since our model relies only on the number of components m, n , the number of pairwise comparisons k , but not the underlying distributions (Lemma 2.1), this paper should be viewed as a nonparametric alternative to current comparison methods. Indeed, one may use this distribution as a performance test between two groups of samples, if nothing is known about the distribution of the samples. Also, one may quantify the increase in reliability of a k -out-of- n system when the number of components n increases. Lastly, we discuss an interesting interpretation of our model, in terms of random walks on \mathbb{Z} .

2. REDUCING TO UNIFORMITY

Denote by $X^{(k)}$ the k -th order statistic of X_1, \dots, X_n , i.e., $X^{(k)}$ is the k th smallest value of $\{X_1, \dots, X_n\}$. The following fundamental lemma simply conveys the exchangeability of c.i.i.d random variables for order statistics.

Lemma 2.1. *Let $X = \{X_1, \dots, X_n\}$ be a set of c.i.i.d. random variables, and let $X^{(i)}$ be the i -th order statistic. For $1 \leq i, j \leq n$, $\Pr[X^{(i)} = X_j] = 1/n$.*

Proof. For a fixed index i , $X^{(i)}$ must be one of X_1, \dots, X_n . By continuity and independence of X_1, \dots, X_n , $\Pr[X_j = X_k] = 0$ over all such pairs. Hence,

$$\Pr[X^{(i)} = X_1] + \dots + \Pr[X^{(i)} = X_n] = 1$$

That X_1, \dots, X_n are identically distributed implies equality between any pair of summands. The lemma follows. \square

Remark 1. The assumption of continuity in Lemma 2.1 is essential. Indeed, $\Pr[X_i = X_j] > 0$ if point masses were present.

Remark 2. Under the assumption of continuity, we may drop the case of equality when comparing X_i 's, since

$$\Pr\left(\bigcup_{i,j}\{X_i = X_j\}\right) = 0.$$

Now, let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be c.i.i.d. random variables, mapping into a totally ordered space $(\Omega, <)$. We seek a probability distribution that compares the bottom k performers from $\mathcal{X} = \{X_1, \dots, X_m\}$, to the bottom k performers from $\mathcal{Y} = \{Y_1, \dots, Y_n\}$; we do so by inducing an order from $(\Omega, <)$.

Definition 1. For random variables X_1, X_2, \dots, X_m , mapping into $(\Omega, <)$, an (ascending) **chain** of length d is the ordering

$$X_{i_1} < X_{i_2} < \dots < X_{i_d}$$

such that, for $1 \leq j \leq d$, $X_{i_j} = X^{(j)}$. Let $B_{m,n}$ be the set of ascending chains of length $m+n$ from the set $\{X_1, \dots, X_m, Y_1, \dots, Y_n\}$ and $C_{m,n}$ the set of ascending chains of length $m+n$ from the set $\{X^{(1)}, \dots, X^{(m)}, Y^{(1)}, \dots, Y^{(n)}\}$.

Obviously, $|B_{m,n}| = (m+n)!$ and $|C_{m,n}| = \binom{m+n}{m}$. When $X_1, \dots, X_m, Y_1, \dots, Y_n$ are c.i.i.d., an easy Corollary to Lemma 2.1 shows the probability distribution on $B_{m,n}$ must also be uniform. As $C_{m,n}$ is the set of equivalence classes in $B_{m,n}$ under the action of $S_m \times S_n$ where the S_m factor permutes X_i 's, and S_n permutes Y_j 's, the probability distribution on $C_{m,n}$ must also be uniform. We record this fact in Lemma 2.2.

Lemma 2.2. *Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be a sequence c.i.i.d. random variables. Then the probability distribution on $C_{m,n}$ is uniform, with*

$$\Pr(c) = \frac{m!n!}{(m+n)!} = \frac{1}{\binom{m+n}{m}}, \quad c \in C_{m,n}$$

Definition 2. Fix fixed k , $1 \leq l \leq k$, denote by $<_{k,l}$ the ordering

$$\mathcal{X} <_{k,l} \mathcal{Y}$$

if and only if $|\{i \mid 1 \leq i \leq k \text{ and } X^{(i)} < Y^{(i)}\}| = l$.

In other words, if $\mathcal{X} <_{k,l} \mathcal{Y}$, then of the k bottom performers, there are exactly l instances when the i -th bottom performer from system X underperformed the i -th bottom performer from system \mathcal{Y} .

Example 1. For $k = 4$, the chain

$$X^{(1)} < Y^{(1)} < Y^{(2)} < X^{(2)} < X^{(3)} < X^{(4)} < X^{(5)} < Y^{(3)} < X^{(6)} < Y^{(4)}$$

in $C_{6,4}$ satisfies

$$X^{(1)} < Y^{(1)}$$

$$X^{(2)} > Y^{(2)}$$

$$X^{(3)} < Y^{(3)}$$

$$X^{(4)} < Y^{(4)}$$

and thus satisfies $\mathcal{X} <_{4,3} \mathcal{Y}$.

As the above example illustrates, for fixed k , each $c \in C_{m,n}$ is canonically associated to an l satisfying $\mathcal{X} <_{k,l} \mathcal{Y}$. By Lemma 2.2, our goal of determining

$$\Pr(\mathcal{X} <_{k,l} \mathcal{Y})$$

amounts to counting those chains in $C_{m,n}$ satisfying $\mathcal{X} <_{k,l} \mathcal{Y}$.

3. COUNTING

We say that a *lattice path* is a path which starts at $(0, 0)$ stepping only in the north $(1, 0)$ and east $(0, 1)$ directions. If we let $\Gamma_{m,n}$ be the set of lattice paths from $(0, 0)$ to (m, n) , then $C_{m,n}$ is in bijective correspondence with $\Gamma_{m,n}$; this can be seen by constructing a lattice path from reading a chain in ascending order, taking a north step when a $Y^{(i)}$ is encountered and an east step for an $X^{(i)}$. Conversely, an ascending chain can be constructed from a lattice path $\gamma \in \Gamma_{m,n}$ by traversing γ in the northeast direction, writing $X^{(i)}$ when the i th horizontal edge is encountered, and writing $Y^{(i)}$ when the i th vertical edge is encountered. Under this bijection, it is clear that

$$|C_{m,n}| = |\Gamma_{m,n}| = \binom{m+n}{m}.$$

Example 2. The chain

$$X^{(1)} < Y^{(1)} < Y^{(2)} < X^{(2)} < X^{(3)} < X^{(4)} < X^{(5)} < Y^{(3)} < X^{(6)} < Y^{(4)}$$

in $C_{6,4}$ corresponds to the path $RUURRRRU$ in $\Gamma_{6,4}$, depicted in figure 1.

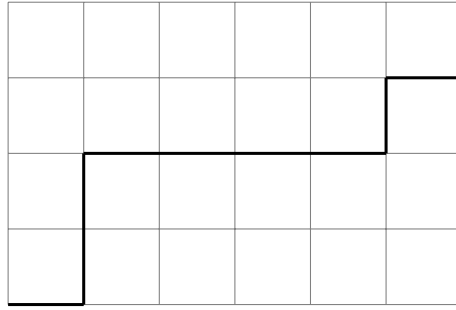


FIGURE 1. $RUURRRRU$ in $\Gamma_{6,4}$.

By lemma 2.2, if $X_1, \dots, X_m, Y_1, \dots, Y_n$ are c.i.i.d., and $\Gamma_{m,n}$ is equipped with the uniform distribution, then the bijection outlined above is also measure preserving. Now, to see what $<_{k,l}$ means in $\Gamma_{m,n}$, we introduce the notion of exceedance.

Definition 3. A path $\gamma \in \Gamma_{m,n}$ is said to have **horizontal exceedance** equal to l iff γ has l horizontal edges below the line $y = x$. Similarly, γ has **vertical exceedance** equal to l iff γ has l vertical edges above the line $y = x$. We say γ has **k -horizontal exceedance** equal to l iff l of the first k horizontal edges lie below the line $y = x$, and γ has **k -vertical exceedance** equal to l iff γ has exactly l of the first k vertical edges lie above the line $y = x$.

The horizontal and vertical exceedance of γ will henceforth be denoted by $HE(\gamma)$ and $VE(\gamma)$ respectively. Similarly, denote by $HE_k(\gamma)$ and $VE_k(\gamma)$ the k -horizontal and k -vertical exceedance, respectively.

Remark 3. When $m = n$, vertical exceedance l paths are typically called (n, l) -flawed Dyck paths in the literature[8].

Example 3. The $\gamma = RUURRRRU$ of example 2 has horizontal exceedance 5 and vertical exceedance 1. If we set $k = 4$, then γ has 4-horizontal exceedance 3, and 4-vertical exceedance 1. The 4-exceedances of γ are depicted in Figure 2.

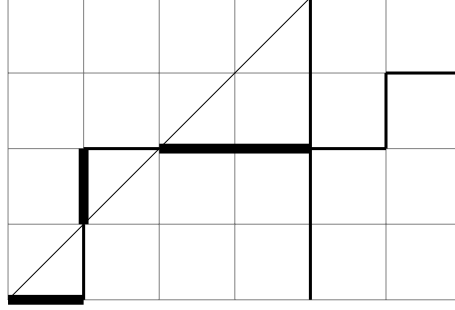


FIGURE 2. $HE_4(\gamma) = 3$, $VE_4(\gamma) = 1$.

Our next lemma relates our method of comparing chains (def. 2) to counting paths in $\Gamma_{m,n}$ of a certain exceedance.

Lemma 3.1. *Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be a sequence of c.i.i.d random variables. Then, for $0 \leq l \leq k \leq m \leq n$,*

$$\Pr(\{c \in C_{m,n} | \mathcal{X} <_{k,l} \mathcal{Y}\}) = \Pr(\{\gamma \in \Gamma_{m,n} | HE_k(\gamma) = l\})$$

Proof. Fix $0 \leq k \leq m \leq n$, and $c \in C_{m,n}$ with corresponding $\gamma \in \Gamma_{m,n}$. Let l be such that

$$\mathcal{X} <_{k,l} \mathcal{Y}$$

is satisfied by c . Then, for exactly l indices between 1 and k , $X^{(j)} < Y^{(j)}$. Equivalently stated, there are l instances when $X^{(j)}$ is appended to an initial chain that contains at least as many $X^{(i)}$'s as $Y^{(i)}$'s. The corresponding statement in $\Gamma_{m,n}$, says there are exactly l instances where the j -th horizontal edge appears before the j -th vertical edge in the initial segment of γ lying in the $k \times k$ box (i.e., the set $[0, k] \times [0, k]$). It is at these indices where a horizontal edge is appended to a path whose initial segment lies on or below the line $y = x$, and this happens exactly l times inside the $k \times k$ box. Thus $HE_k(\gamma) = l$. \square

Corollary 3.2. *Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be a sequence of c.i.i.d. random variables. Then, for fixed k ,*

$$\Pr(\{c \in C_{n,m} | \mathcal{X} >_{k,l} \mathcal{Y}\}) = \Pr(\{\gamma \in \Gamma_{m,n} | VE_k(\gamma) = l\})$$

Proof. Follows by reflecting through the line $y = x$, and applying Lemma 3.1. \square

By Lemma 2.1, and the fact that $|C_{m,n}| = |\Gamma_{m,n}| (= \binom{m+n}{m})$, finding $|\{c \in C_{n,m} | \mathcal{X} >_{k,l} \mathcal{Y}\}|$ is equivalent to counting the number of elements in $\{\gamma \in \Gamma_{m,n} | VE_k(\gamma) = l\}$. We count the latter set by first producing a closed formula for the number of paths with horizontal exceedance equal to l that terminate at a lattice point (x, y) . We then outline an enumeration scheme to count k -horizontal exceedant paths in $\Gamma_{m,n}$.

To carry out the first step, write $\#(x, y, l)$ for the number of lattice paths η that terminate at the point (x, y) , satisfying $HE(\eta) = l$. Clearly, $\#(x, y, l) = 0$ if $l > x$. When $x = y$, a

celebrated theorem of Chung and Feller [4] states that lattice paths terminating at (i, i) are partitioned evenly among all possible exceedances. In particular, $\#(i, i, l)$ is the i th Catalan number C_i . The next lemma simply combines the previous two facts.

Theorem 3.3 (Chung-Feller).

$$(3.1) \quad \#(i, i, l) = \begin{cases} C_i = \frac{1}{i+1} \binom{2i}{i} & 0 \leq i \leq l \\ 0 & \text{otherwise} \end{cases}$$

A beautiful bijective proof of the Chung-Feller Theorem can be found in [6].

The following Lemma and Corollary reduces the task of computing $\#(x, y, l)$ to the case when $y < x$.

Lemma 3.4. For $\gamma \in \Gamma_{x,y}$ and $0 \leq k \leq \min\{x, y\}$,

$$(3.2) \quad HE_k(\gamma) + VE_k(\gamma) = k.$$

Proof. Observe, k -exceedance of $\gamma \in \Gamma_{x,y}$ is completely determined by its initial segment having only one vertex, the terminal vertex, lying on the boundary of the $k \times k$ box. Such an initial segment has a canonical extension to a path in $\Gamma_{k,k}$ by appending edges from the terminal vertex to (k, k) , and this extension does not affect exceedances. So, if $HE_k(\gamma) = l$, then there is a unique path $\tilde{\gamma}$ in $\Gamma_{k,k}$ that agrees with γ 's aforementioned initial segment, making $HE_k(\tilde{\gamma}) = l$ as well. Similarly $VE_k(\gamma) = VE_k(\tilde{\gamma})$. By monotonicity, $\tilde{\gamma}$ has $2l$ edges below the line $y = x$, and each path in $\Gamma_{k,k}$ has exactly $2k$ edges, so $\tilde{\gamma}$ has $2(k - l)$ edges above $y = x$, thus $VE_k(\tilde{\gamma}) = k - l$, and

$$HE_k(\gamma) + VE_k(\gamma) = HE_k(\tilde{\gamma}) + VE_k(\tilde{\gamma}) = k$$

The lemma is proved. □

Corollary 3.5. For $\gamma \in \Gamma_{x,y}$.

$$(3.3) \quad HE(\gamma) + VE(\gamma) = \max\{x, y\}.$$

Proof. By noting that $\gamma \in \Gamma_{x,y}$ extends uniquely to a path in $\Gamma_{\max\{x,y\}, \max\{x,y\}}$ without affecting exceedances, and setting $k = \max\{x, y\}$, the Corollary follows from Lemma 3.4. □

By reflecting through the line $y = x$, Corollary 3.5 yields the identity

$$(3.4) \quad \#(x, y, l) = \#(y, x, \max\{x, y\} - l)$$

so we need only to determine $\#(x, y, l)$ for the case $y < x$. To do so, we invoke another classical result, due to Bertrand [3], which states there are

$$\binom{y+x}{y} - \binom{y+x}{y-1}$$

lattice paths from $(0, 0)$ to (x, y) , $x > y$, that never go above the line $y = x$ (so contact at $y = x$ is allowed). An easy extension ([1],[2]) shows that there are

$$(3.5) \quad \binom{x+y-1}{y} - \binom{x+y-1}{y-1} = \frac{x-y}{x+y} \binom{x+y}{y}$$

lattice paths from $(0, 0)$ to (x, y) that stay strictly below the line $y = x$, except at $(0, 0)$, known as the *ballot number* $b(x, y)$.

Theorem 3.6. For $x > y$, If l satisfies $x - y \leq l \leq x$,

$$(3.6) \quad \#(x, y, l) = \sum_{i=x-l}^y C_i b(x-i, y-i)$$

and $\#(x, y, l) = 0$ for l otherwise.

Proof. Fix (x, y) and l . We partition $A = \{\eta \in \Gamma_{x,y} | HE(\eta) = l\}$ into the (possibly empty) sets A_0, A_1, \dots, A_y , where A_i is the set of lattice paths η that last cross the line $y = x$ at the point (i, i) . To ensure A is nonempty, l is at most x , and since a lattice path to (x, y) has horizontal exceedance at least $x - y$,

$$\#(x, y, l) = 0, \text{ if } x > l \text{ or } l < x - y$$

Because $x > y$, the terminal segment of $\eta \in A_i$, i.e., the segment of η from (i, i) to (x, y) , must lie below the line $y = x$, by the “last cross” property of A_i . Thus, the terminal segment of $\eta \in A_i$ will always contribute $(x - i)$ to the horizontal exceedance, mandating a horizontal exceedance of $l - (x - i)$ in η ’s initial segment. Thus, to ensure A_i is nonempty,

$$l - (x - i) \geq 0, \text{ or } i \geq x - l$$

So nonempty A_i run from $i = x - l$ to $i = y$. We call such indices admissible. To count nonempty A_i , we multiply the number of initial segments by the number of terminal segments as they were described above. By Lemma 3.3, the number initial segments terminating at (i, i) of horizontal exceedance $l - (x - i)$ is

$$C_i = \frac{1}{i+1} \binom{2i}{i}.$$

Applying equation 3.5, the number of terminal (monotonic) paths from (i, i) to (x, y) that stay below the line $y = x$ is given by

$$b(x-i, y-i) = \frac{x-y}{x+y-2i} \binom{x+y-2i}{y-i},$$

making

$$|A_i| = C_i b(x-i, y-i).$$

Summing over admissible i completes the theorem. \square

Remark 4. By deleting the terminal edge, one can see equation 3.6 satisfies the necessary recursion relation

$$\#(x, y, l) = \#(x-1, y, l-1) + \#(x, y-1, l)$$

at those points (x, y) with $x > y$.

Denote by $\#(m, n, k, l)$ the number of lattice paths terminating at (m, n) having k -horizontal exceedance l . We now prove the main theorem.

Theorem 3.7.

$$(3.7) \quad \#(m, n, k, l) = \sum_{j=0}^{k-1} I(\{j \geq l\}) |B_{(j,k)}(l)| + \sum_{j=0}^{k-1} I(\{j \geq k-l\}) |B_{(k,j)}(l)|$$

where I is the indicator function,

$$\begin{aligned}
|B_{(k,j)}(l)| &= \binom{m+n-k-j}{m-k} \sum_{i=k-l}^j C_i b(k-1-i, j-i), \quad 0 \leq j < k-1 \\
|B_{(j,k)}(l)| &= \binom{m+n-k-j}{n-k} \sum_{i=l}^j C_i b(k-1-i, j-i), \quad 0 \leq i < k-1 \\
|B_{(k-1,k)}(l)| &= \binom{m+n-2k+1}{n-k} C_{k-1} \\
|B_{(k,k-1)}(l)| &= \binom{m+n-2k+1}{m-k} C_{k-1}
\end{aligned}$$

Proof. The proof will have a similar flavor to the proof of Theorem 3.6. Fix m, n, k , and l . As stated before, k -horizontal exceedance of $\gamma \in \Gamma_{m,n}$ is determined once its initial segment first reaches either of the lines $x = k$, $y = k$ first, whence we invoke Theorem 3.6. So, we partition $B = \{\gamma \in \Gamma_{m,n} | HE_k(\gamma) = l\}$ into the (possibly empty) sets

$$B_{(0,k)}, B_{(1,k)}, \dots, B_{(k-1,k)}, B_{(k,k-1)}, \dots, B_{(k,1)}, B_{(k,0)},$$

where $B_{(j,k)}$ (resp. $B_{(k,j)}$) is the set of $\gamma \in \Gamma_{(m,n)}$ with $HE_k(\gamma) = l$, possessing an initial segment to first cross the line $y = k$ (resp. $x = k$) at the point (j, k) (resp. (k, j)). We will work first with $B_{(k,j)}$. Notice, for a lattice path $\gamma \in B_{(k,j)}$, the initial segment that terminates at the vertex (k, j) determines k -horizontal exceedance. As $j < k$, this initial segment has its last edge below the line $y = x$ and by definition of $B_{(k,j)}$, this edge must be horizontal. Since this last edge contributes 1 to the horizontal exceedance, $\#(k-1, j, l-1)$ counts the number of such initial segments in $B_{(k,j)}$. This argument holds except at $j = k-1$, where $\#(k-1, k-1, l-1)$ is C_{k-1} , by Lemma 3.3. Once k -exceedance is determined, we may append to the initial segment any of the

$$\binom{m+n-k-j}{m-k}$$

lattice paths from (k, j) to (m, n) , constructing a path in $B_{(k,j)}$. Multiplying, we get, for j satisfying $k-l \leq j < k-1$,

$$\begin{aligned}
|B_{(k,j)}| &= \#(k-1, j, l-1) \binom{m+n-k-j}{m-k} \\
&= \binom{m+n-k-j}{m-k} \sum_{i=k-l}^j C_i b(k-1-i, j-i), \text{ and} \\
|B_{(k,k-1)}| &= \binom{m+n-2k+1}{m-k} C_{k-1}
\end{aligned}$$

Counting paths in $B_{(j,k)}$ is slightly different, since the initial segment that determines k -horizontal exceedance has its last edge *above* the line $y = x$, and this last edge is vertical. Since this edge doesn't contribute to horizontal exceedance, $\#(j, k-1, l)$ counts the number of such initial segments in $B_{(j,k)}$. As before, we may append to this initial segment any of

the

$$\binom{m+n-k-j}{n-k}$$

lattice paths from (j, k) to (m, n) , to construct a path in $B_{(j, k)}$. Recall the reflection identity (Eq. 3.4)

$$\#(j, k-1, l) = \#(k-1, j, k-1-l).$$

Thus, for $l \leq j < k-1$,

$$\begin{aligned} |B_{(j, k)}| &= \#(k-1, j, k-1-l) \binom{m+n-k-j}{n-k} \\ &= \binom{m+n-k-j}{n-k} \sum_{i=l}^j C_i b(k-1-i, j-i), \text{ and} \\ |B_{(k-1, k)}| &= \binom{m+n-2k+1}{n-k} C_{k-1} \end{aligned}$$

It remains to show which of the sets

$$B_{(0, k)}, B_{(1, k)}, \dots, B_{(k-1, k)}, B_{(k, k-1)}, \dots, B_{(k, 1)}, B_{(k, 0)},$$

are nonempty, for our l .

If $l = 0$, no horizontal lines lie below $y = x$ in the $k \times k$ box, so the nonempty sets are

$$B_{(0, k)}, \dots, B_{(k-1, k)}.$$

Summing, we get

$$\#(m, n, k, 0) = \left(\sum_{j=0}^{k-1} |B_{(j, k)}| \right)$$

If $l = k$, all horizontal lines lie below $y = x$ in the $k \times k$ box, so the nonempty sets are

$$B_{(k, 0)}, \dots, B_{(k, k-1)}.$$

Summing, we get

$$\#(m, n, k, k) = \left(\sum_{j=0}^{k-1} |B_{(k, j)}| \right)$$

Case $0 < l < k$: $HE_k(\gamma) = l$ admits nonemptiness in only the sets

$$B_{(k, k-l)}, \dots, B_{(k, k-1)}, B_{(k-1, k)}, \dots, B_{(l, k)}.$$

Summing, we get

$$\#(m, n, k, l) = \sum_{j=k-l}^{k-1} |B_{(k, j)}| + \sum_{j=l}^{k-1} |B_{(j, k)}|$$

All cases align precisely with formula (3.7), as stated. □

Remark 5. Theorem 3.7 provides the probability distribution in question. That is

$$(3.8) \quad \Pr(\mathcal{X} <_{k,l} \mathcal{Y}) = \frac{\#(m, n, k, l)}{\binom{m+n}{m}}$$

4. RANDOM WALKS

Lemma 3.1 established a bijection between the sets

$$\{c \in C_{m,n} | \mathcal{X} <_{k,l} \mathcal{Y}\} \text{ ,and } \{\gamma \in \Gamma_{m,n} | HE_k(\gamma) = l\}.$$

The mapping

$$(1, 0) \rightarrow +1$$

$$(0, 1) \rightarrow -1$$

effortlessly establishes a bijection between the set $\Gamma_{m,n}$ and the set $W_{m,n}$ of length $m+n$ integer walks on \mathbb{Z} that start at $x=0$ and terminate at $x=m-n$. As mentioned before, by monotonicity in $\Gamma_{m,n}$, any γ that satisfies $HE_k(\gamma) = l$, has exactly $2l$ of its first $2k$ steps lying below the main diagonal. In $W_{m,n}$, these paths spend exactly $2l$ of their first $2k$ steps above $x=0$. Thus, if we let $T_{2k}: W_{m,n} \rightarrow \{0, 2, \dots, 2k\}$ be the number of initial $2k$ steps lying above $x=0$, we have yet another interpretation of $\mathcal{X} <_{k,l} \mathcal{Y}$:

$$\Pr(\{c \in C_{m,n} | \mathcal{X} <_{k,l} \mathcal{Y}\}) = \Pr(\{w \in W_{m,n} | T_{2k}(w) = 2l\})$$

5. FUTURE RESEARCH

It is desirable to understand the asymptotics of the distribution in eq. 3.8. Heuristic evidence suggests that

$$(5.1) \quad \frac{k}{\binom{m+n}{m}} \#(m, n, k, xk), x \in \{0, 1/k, 2/k, \dots, 1\}$$

is a discretized beta distribution, where the two shape parameters depend, obviously, on m, n and k . The nature of this dependency, however, could not be found. A particularly intriguing case is a symmetric one, $n = 2k, m = 2k$, where it appears the distribution in eq. 5.1 converges, in probability, to an arcsine-like distribution as $k \rightarrow \infty$. This would have a remarkable implication in how our distribution relates, in terms of random walks, to the celebrated arcsine law of Lévy.

REFERENCES

- [1] D. Andre. Calcul des probabilités. Solution direct du probleme resolu par M. Bertrand. *Comptes Rendus de l'Academie des Sciences*, 105:436, 1887.
- [2] E. Barbier. Calcul des probabilités. Generalization du probleme resolu par M.J.Bertrand. *Comptes Rendus de l'Academie des Sciences*, 105:407, 1887.
- [3] J. Bertrand. Calcul des probabilités. Solution d'un probleme. *Comptes Rendus de l'Academie des Sciences*, 105:369, 1887.
- [4] K. Chung and K. Feller. On fluctuations in coin tossing. *Proc. Nat. Acad. Sci. USA*, 35:605–608, 1949.
- [5] R. Dykstra, S. Kochar, and J. Rojo. Stochastic comparisons of parallel systems of heterogeneous exponential components. *J. Statist. Plann. Inference*, 65:203–211, 1997.
- [6] I. Goulden and L. Serrano. Maintaining the spirit of the reflection principle when the boundary has arbitrary integer slope. *J. Combin. Theory Ser A*, 104:317–326, 2003.
- [7] B. Khaleli and S. Kochar. Some new results on stochastic comparisons of parallel systems. *J. Appl. Prob.*, 37:1123–1128, 2000.

- [8] J. Ma and Y.Yeh. Generalizations of Chung-Feller theorems. *Bull. Inst. Math. Acad. Sin. (N.S.)*, 4(3):299–332, 2009.